SWAMI RAMANAND TEERTH MARATHWADA UNIVERSITY, NANDED
MAHARASHTRA, INDIA
(*UGC recognized, NAAC Re-accredited 'B++' grade state university*)

"PREDICTION OF SOME COMMODITIES IN INDIA - A STATISTICAL DATA MINING APPROACH TO EVALUATION"

Research Proposal submitted in Partial fulfillment of registration to Ph.D Program

In the Subject

Statistics faculty of science and technology

Submitted by

MISS. PAYAL AVINASH MAHAJAN
M.Sc. (Statistics)
No:7350707377
Email ID: payalmahajan81298@gmail.com

Under The Supervision Of,
DR. MADHAV RAMRAO FEGADE
Assistant Professor& Head
Department of Statistics
Digambarrao Bindu Arts, Commerce and Science College Bhokar, Dist. Nanded, Maharashtra, India - 431801.
Phone No:9922675834

Name of Research Center

School of Mathematical Science, S.R.T.M.University, Nanded.431606

2024

INDEX

Sr. No.	Contents	Page No.
1	Introduction	3
2	Review of Literature and Development in the subject	5
3	Objectives of the research	6
4	Methodology	6
5	Importance of study	7
6	Proposed work plan	8
7	References	9

1. INTRODUCTION

Data mining is the process of analyzing data from different perspectives and summarizing it into usefulinformation. Data mining is one of the analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. Data mining methods analyses relationships and patterns in stored data. Data mining techniques tends to identify interesting patterns among the dataset. Data mining is a technology which is used to find interesting pattern between huge datasets. [14]

Commodity market is said to be a huge collection of various commodities, Gold, Oil, etc, which are referred to as hard commodities. In ancient days, gold coins were a medium of exchange. Another important commodity is oil. The price of oil changes daily, which has an impact on every goods and services provided. A country can make a payment via paper currency. This mode can be changed to exchange of gold at a fixed rate. The exchange rate between currencies was based on the amount of currencies needed to purchase one ounce of gold. US dollar(USD) is widely accepted as an instrument of global currency exchange. The gold price is directly related to USD. [14] Examines the relationship between the rate of gold and oil with respect to USD. This also explores the commodity Market based on USD. The price of Gold, Oil and the US dollar share different relationships, in different circumstances and that explores the interesting pattern existed in the commodity market.

Data mining in the commodity market refers to the process of discovering valuable insights, patterns, and relationships from large and complex datasets related to commodity trading, prices, supply and demand, and other relevant factors. It involves using advanced computational and statistical techniques to extract meaningful information that can aid in decision-making, risk management, and strategy formulation within the context of commodity trading and investment.

- **Commodity Market**

A commodity market is a financial marketplace where raw materials or primary agricultural products are traded. These commodities are standardized and interchangeable with other goods of the same type. The key characteristic of commodity markets is that they facilitate the buying and selling of these raw materials or primary products, often referred to as "commodities," based on prevailing market prices.

Commodity markets serve several essential functions in the global economy:

- Price Discovery: Commodity markets provide a platform for buyers and sellers to interact and determine the market-clearing price for a specific commodity. These prices are influenced by various factors such as supply and demand dynamics, geopolitical events, weather conditions, and global economic trends.[22]
- Risk Management: Commodity markets offer tools for hedging against price volatility. Participants, such as producers, consumers, and investors, use futures contracts and options to lock in prices for future delivery, thereby reducing their exposure to price fluctuations.[23]
- Investment Opportunities: Commodity markets attract investors seeking to diversify their portfolios. Commodities often have low correlation with traditional financial assets like stocks and bonds, making them attractive for risk management and potential returns.[22]
- Supply and Demand Balancing: Commodities play a crucial role in ensuring a balance between global supply and demand. If demand for a certain commodity increases, its price rises, incentivizing producers to increase supply. Conversely, if demand decreases, prices may fall, potentially leading to decreased production.[23]

- Global Trade: Many commodities are traded on international exchanges, enabling countries to participate in the global economy by exporting their surplus production and importing items they lack.
- Impact on Inflation: Changes in commodity prices can have a significant impact on inflation rates. Rapid increases in commodity prices can lead to cost-push inflation, affecting the prices of a wide range of goods and services.[23]

Commodity markets are often categorized into two main types:

- Hard Commodities: These include tangible goods extracted from the earth, such as metals (gold, silver, copper), energy resources (crude oil, natural gas), and agricultural products (wheat, corn, soybeans).
- Soft Commodities: These are typically agricultural products and other renewable resources, such as coffee, cocoa, cotton, sugar, and livestock.

Commodity markets can be physical markets, where actual delivery of the commodity occurs, or futures markets, where contracts for future delivery are bought and sold. The latter is particularly common in modern commodity trading. Overall, commodity markets are essential components of the global economy, contributing to price stability, risk management, investment opportunities, and international trade.[22]

2. REVIEW OF LITERATURE AND DEVELOPMENT IN SUBJECT

Ben et al.[2]focused on the aspect of forecasting commodity prices, as they play an important role in providing implications for economic situations. They showed the main mechanisms that can cause commodity prices affects the economy and financial systems. Particularly, they discussed the theoretical motivation regarding why both developing and developed countries show a great interest in following the behaviour and pattern of the main commodity markets. Therefore, the forecasting of commodity prices is a vital activity for investors, hedgers, and policy makers. Their study proposed the use of various DL tools and showed that the LSTM method is the best-fit model. They highlight the superiority of Bloomberg Livestock Sub index and Bloomberg Industrial Metals Sub index compared to other commodity indices. Magner et al. [11] studied documents that commodity currencies have a predictive ability on the synchronization behaviour of commodity market returns, considering base metals, precious metals, and energy commodities. Findings of their study support the commodity currency hypothesis and show yet another instance in which currencies of major commodity exporting economies can predict changes in the aspect of commodity prices. They expand the use of asset networks, which have been widely applied in stock markets, to analyze the synchronization phenomenon of commodities. Their methodology provides an important empirical reference point for the application of network analysis methods and forecasting techniques to study the phenomenon of commodity market synchronization of returns and its impact on systemic risk.

Jesus et al. [5] presented a model to estimate the price of commodities in the Brazilian market from short-term timeseries forecasting. In their research, two databases were considered, one referring to the abroad market (opening values, max, min, closing adjustment, and volume) and the Brazilian market (the price of the day), considering several different commodities. Subsequently, the technical indicators were evaluated from the TA-Lib. Pearson's correlation coefficient was used, and records with low correlation were not used. Hassani et al. [8**]** showed that review of published work to date on the application ofdata mining in official statistics, and on identification of the techniques that have been explored. In addition, the importance of data mining to official statistics is flagged and a

summary of the challenges that have hindered its development over the course of the last two decades is presented.

Yahia et al. [19] tried to put a new direction for the evaluation of some techniques for solving data mining tasks such as: Statistics, Visualization, Clustering, Decision Trees, Association Rules and Neural Networks. The new approach has succeed in defining some new criteria for the evaluation process, and it has obtained valuable results based on what the technique is, the environment of using each techniques, the advantages and disadvantages of each technique, the consequences of choosing any of these techniques to extract hidden predictive information from large databases, and the methods of implementation of each technique. Finally, their work presented some valuable recommendations in this field. Bowman et al. [4] forecasted **t**he performance of commodity price based on judgment, historical price data, and futures of commodities is compared with historical-data-based forecasts, indicating that futures-based models yield better forecasts.The share of primary commodities in global output and trade has declined over the past century. But fluctuations in commodity prices continue to affect global economic activity. Their results of their study reveal that futures prices can provide reasonable guidance about likely developments in spot prices. Futures prices should be unbiased predictors of future spot prices. Sathya et al. [17] used some pre-processing techniques and finally plotted MSP, actual and de-seasonalized prices of the crop. Outlier Detection through Inter-quartile ranges, Statistical decomposition of Time-series data, Data visualization is performed for the different APMCs of Maharashtra. Runqiua et al. [16] predicted gas emission in coal mines, the complicated nonlinear characteristics of gas emission was analysis; the prediction method was put forward for gas emission based on self-organizing data mining. It was used the ternary quadratic polynomial for the local function and the original variable was used in each generation, and the minimum deviation principle was used for criteria of model selected. Afterwards, the high-order equation of prediction was established for gas emission by self-organizing data mining method.

Mulik et al. [12] deliberated use of various data mining methods and tools in agriculture business sector. It is used it to solve yield estimation of crop. Crop yield estimation needs to be crop planning maximize crop production & food security. This research will guide farmers to select correct crops in particular seasons. So crop production will increase and automatically effects on agribusiness. At the country level, yield forecasting is used in the determination of national food security, crop insurance policy, import and export plans, and government aid for farmers. Yield estimation means predicting the crop productivity. Generating rules with higher accuracy for agriculture databases can be done using different methods of data mining.

3. OBJECTIVES

- To explore some commodities from various sector in India. (Viz. Agriculture sector, Industrial Sector, etc.)
- To study various commodities that makes higher impact on economical growth of India.
- To study and analyse agriculture commodities and their seasonal aspects.
- To study the time series pattern & variability of different commodities.
- Try to develop some classification models for commodities.
- Try to forcast commodity prices for next few years based on the praposed model.

4. METHODOLOGY

In this study, the data will be gathered from various sources viz. different gathered websites, government offices for our study. This may include historical price data, production and consumption

figures, weather data, economic indicators, geopolitical events, and any other variables that might influence commodity prices. The details are as follows:

1. Identify Commodities: Determine the specific commodities you want to predict. Commodities can range from agricultural products like wheat, rice, and cotton to energy resources like crude oil, natural gas, and metals like gold, silver, and copper, etc.

2. Sources of Data: Identify credible and reputable sources of data for the selected commodities. Some potential sources include government agencies, international organizations, commodity exchanges, industry reports, research publications, and financial news outlets. Here are a few examples of data sources:

- **Government Agencies:** Ministries of agriculture, commerce, and finance often provide data on production, consumption, imports, and exports of commodities.
- **Commodity Exchanges:** Websites of commodity exchanges like MCX (Multi Commodity Exchange) in India provide historical price data for various commodities.
- **International Organizations:** Organisations like the World Bank, United Nations, and International Monetary Fund (IMF) provide comprehensive global economic data including commodity-related information.
- **Industry Reports:** Research firms and industry associations often publish reports on commodity market trends, supply-demand dynamics, and price forecasts.
- **Financial News:** Financial news websites and publications frequently provide insights into factors affecting commodity prices.

.

5. IMPORTANCE OF THE STUDY

Studying the commodity market holds significant importance due to its impact on various aspects of the global economy, investment strategies, and supply chain dynamics. The importance of studying the commodity market is classified as follows:

- **National Status** :

Data mining plays an important role in financial fraud ditection, as it is often applied to extract and uncover the hidden truths behind very large quantities of data. Turban et al. [18] defined data mining as a process that uses statistical, mathematical, artificial intelligence, and machine-learning techniques to extract and identify useful information and subsequently gain knowledge from a large database. Frawley et al. [7] stated that the objective of data mining is to obtain useful, non-explicit information from data stored in large repositories. Kou et al. [9] highlighted that an important advantage of data mining is that it can be used to develop a new class of models to identify new attacks before they can be detected by human experts. Phua et al. [15] pointed out that fraud detection has become one of the best established applications of data mining in both industry and government. Knowledge discovery in databases (KDD) has become a process of considerable interest in recent years as the amounts of data in many databases have grown tremendouslylarge. KDD means the application of non-trivial procedures for identifyingeffective, coherent, potentially useful, and previously unknown patterns in large databases. The KDD process generally consists of pre-processing, data mining and post-processing. Due to the importance of data mining to KDD, many researchers in database and machine learning fields are primarily interested in this new research topic because it offers opportunities to discovering useful information and important relevant patterns in large databases, thus helping decision-makers easily analyze the data and make good decisions regarding the domains concerned.[15]

Over the past few years, a number of review articles have appeared in conferences or journal publications. However, their survey presented herein is an up-to-date, comprehensive and state-of-the-art review of data mining applications in FFD.

The data mining approach has both national and international scopes, and it can be applied in various domains and contexts to extract valuable insights from large datasets. Here's an overview of how the data mining approach can be applied at both levels.[24]

- **Scope of study**:

Global Health: International health organizations can collaborate using data mining to analyse health data from various countries, enabling the identification of global health trends, early detection of pandemics, and facilitating the sharing of best practices in healthcare.Climate Change: Data mining can help analyze climate data from around the world, assisting researchers and policymakers in understanding the impacts of climate change, predicting weather patterns, and formulating strategies for mitigation and adaptation. International Trade and Commerce: Data mining can aid in analyzing international trade patterns, identifying emerging markets, predicting trade disruptions, and assisting policymakers in making informed decisions related to trade agreements. Statistical methods used to detecting fraud, including credit card fraud, money laundering, telecommunications fraud, etc [3], also use to survey financial applications of data mining including stock market and bankruptcy predictions and fraud detection [21], survey of data mining-based fraud detection research, including credit transaction fraud, telecoms subscription fraud, automobile insurance fraud and the viz., [15] insurance fraud [6] and financial statement fraud [20]. Security and Terrorism Analysis: International security agencies can employ data mining to analyze global security threats, track patterns of terrorism and criminal activity, and collaborate on strategies to counteract these threats. Social and Cultural Studies: Researchers can use data mining to analyze social media trends, global cultural preferences, and communication patterns to gain insights into the evolving global social landscape. Environmental Conservation: International organizations can employ data mining to analyze environmental data from various regions, helping to identify endangered species, monitor deforestation, and implement conservation efforts on a global scale.

The quantitative method used to highlight dependencies in the dataset. The method Bayesian confidence propagation neural network (BCPNN) issued to highlight dependencies in the data set. The method uses Bayesian statistics implemented in neural network architecture to analyse all reported drug adverse reaction combinations. This method is now in routine use for drug adverse reaction signal detection. Also this approach has been extended to highlight drug group effects and look for higher order dependencies in the WHO data. [1]

6. PROPOSED WORK PLAN

Tentative year wise plan of work & targets to be achieved during study:

Year 1:

- To identify the specific commodities based on the national importance and availability of the datasets on various platforms.
- To collect and study of structure of commodities.
- Study of different data mining packages in different software.
- Collect secondary data from different websites of government of India.

Year 2:

- Try to develop new model to evaluate predication of some commodities in India. (viz. crop, gas, metal, coal, etc)
- Apply statistical tools or software to summarize and analyse the data.

Year 3:

- Publication of the findings of the research.
- Publication of papers in various national and international journals and conferences.
- Thesis writing.

7. REFERANCES

1. Bate, Andrew, Marie Lindquist, I. Ralph Edwards, and Roland Orre. "A data mining approach for signal detection and analysis." *Drug safety* 25 (2002): 393-397.
2. Ben Ameur, Hachmi, Sahbi Boubaker, ZiedFtiti, Wael Louhichi, and Kais Tissaoui. "Forecasting commodity prices: empirical evidence using deep learning tools." *Annals of Operations Research* (2023): 1-19.
3. Bolton, Richard J., and David J. Hand. "Statistical fraud detection: A review." *Statistical science* 17, no. 3 (2002): 235-255.
4. Bowman, Chakriya, and Aasim M. Husain. *Forecasting commodity prices: Futures versus judgment*. March, 2004.
5. de Jesus, Gabriel Schneider, Aurélio Faustino Hoppe, and Andreza Sartori. "Time Series Forecasting Using LSTM Applied for Commodity Futures Price in Latin America." *Available at SSRN 4352469* (2023).
6. Derrig, Richard A. "Insurance fraud." *Journal of Risk and Insurance* 69, no. 3 (2002): 271-287.
7. Frawley, William J., Gregory Piatetsky-Shapiro, and Christopher J. Matheus. "Knowledge discovery in databases: An overview." *AI magazine* 13, no. 3 (1992): 57-57.
8. Hassani, Hossein, Gilbert Saporta, and Emmanuel Sirimal Silva. "Data mining and official statistics." *databases* 6 (2014): 76.
9. Kou, Yufeng, Chang-Tien Lu, Sirirat Sirwongwattana, and Yo-Ping Huang. "Survey of fraud detection techniques." In *IEEE International Conference on Networking, Sensing and Control, 2004*, vol. 2, pp. 749-754. IEEE, 2004.
10. Lee, Yeong-Chyi, Tzung-Pei Hong, and Wen-Yang Lin. "Mining association rules with multiple minimum supports using maximum constraints." *International Journal of Approximate Reasoning* 40, no. 1-2 (2005): 44-54.

11. Magner, Nicolas S., Nicolás Hardy, Jaime Lavin, and Tiago Ferreira. "Forecasting Commodity Market Synchronization with Commodity Currencies: A Network-Based Approach." Entropy 25, no. 4 (2023): 562..

12. Mulik, Sujata, and Ajit More. "future of agriculture with data mining." Shodha Prabha (UGC CARE Journal) Vol. 47, Issue. 03, No.3: 2022

13. Ngai, Eric WT, Yong Hu, Yiu Hing Wong, Yijun Chen, and Xin Sun. "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature." Decision support systems 50, no. 3 (2011): 559-569.

14. Parkavi, S., S. Sasikumar, and M. Venkatesh Saravanakumar. "Prediction of Commodities Market by Using Data Mining Technique." i-Manager's Journal on Computer Science 4, no. 1 (2016): 16.

15. Phua, Clifton, Vincent Lee, Kate Smith, and Ross Gayler. "A comprehensive survey of data mining-based fraud detection research." arXiv preprint arXiv:1009.6119 (2010).

16. Runqiu, Li, Shi Shiliang, Wu Aiyou, Luo Wenke, and Zhu Hongping. "Research on prediction of gas emission based on self-organizing data mining in coal mines." Procedia Engineering 84 (2014): 779-785.

17.Sathya, K., and R. Karthiban. "Time Series Analysis on Agricultural Commodity Prices." Asian Journal of Applied Science and Technology (AJAST) 5, no. 2 (2021): 36-41.

18. Turban, Efraim. Decision support and business intelligence systems. Pearson Education India, 2011.

19. Yahia, MoawiaElfaki, and Murtada El-mukashfi El-taher. "A new approach for evaluation of data mining techniques." International Journal of computer Science issues 7, no. 5 (2010): 181-186.

20. Yeh, I-Cheng, and Che-hui Lien. "The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients." Expert systems with applications 36, no. 2 (2009): 2473-2480.

21.Zhang. Dongsong, and Lina Zhou. "Discovering golden nuggets: data mining in financial application." IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 34, no. 4 (2004): 513-522.

22. Chandra, R. (2013). Financial management. BookRix.

23. Ranganatham, M. (2006). Investment analysis and portfolio management. Pearson Education India.

24. Berzal, F., & Matín, N. (2002). Data mining: concepts and techniques by Jiawei Han and Micheline Kamber. ACM Sigmod Record, *31*(2), 66-68.

Signatur	Signatur
Research Scholar	Research Guide
Miss. Payal Avinash Mahajan	Dr. Madhav R. Fegade